



**University of  
Zurich<sup>UZH</sup>**

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2014

---

## **Music and language expertise influence the categorization of speech and musical sounds: behavioral and electrophysiological measurements**

Elmer, Stefan ; Klein, Carina ; Kühnis, Jürg ; Liem, Franziskus ; Meyer, Martin ; Jäncke, Lutz

**Abstract:** In this study, we used high-density EEG to evaluate whether speech and music expertise has an influence on the categorization of expertise-related and unrelated sounds. With this purpose in mind, we compared the categorization of speech, music, and neutral sounds between professional musicians, simultaneous interpreters (SIs), and controls in response to morphed speech-noise, music-noise, and speech-music continua. Our hypothesis was that music and language expertise will strengthen the memory representations of prototypical sounds, which act as a perceptual magnet for morphed variants. This means that the prototype would "attract" variants. This so-called magnet effect should be manifested by an increased assignment of morphed items to the trained category by a reduced maximal slope of the psychometric function as well as by differential event-related brain responses reflecting memory comparison processes (i.e., N400 and P600 responses). As a main result, we provide first evidence for a domain-specific behavioral bias of musicians and SIs toward the trained categories, namely music and speech. In addition, SIs showed a bias toward musical items, indicating that interpreting training has a generic influence on the cognitive representation of spectrotemporal signals with similar acoustic properties to speech sounds. Notably, EEG measurements revealed clear distinct N400 and P600 responses to both prototypical and ambiguous items between the three groups at anterior, central, and posterior scalp sites. These differential N400 and P600 responses represent synchronous activity occurring across widely distributed brain networks and indicate a dynamical recruitment of memory processes that vary as a function of training and expertise.

DOI: [https://doi.org/10.1162/jocn\\_a\\_00632](https://doi.org/10.1162/jocn_a_00632)

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-95986>

Journal Article

Originally published at:

Elmer, Stefan; Klein, Carina; Kühnis, Jürg; Liem, Franziskus; Meyer, Martin; Jäncke, Lutz (2014). Music and language expertise influence the categorization of speech and musical sounds: behavioral and electrophysiological measurements. *Journal of Cognitive Neuroscience*, 26(10):2356-2369.

DOI: [https://doi.org/10.1162/jocn\\_a\\_00632](https://doi.org/10.1162/jocn_a_00632)

# Music and Language Expertise Influence the Categorization of Speech and Musical Sounds: Behavioral and Electrophysiological Measurements

Stefan Elmer<sup>1\*</sup>, Carina Klein<sup>1\*</sup>, Jürg Kühnis<sup>1</sup>, Franziskus Liem<sup>1</sup>,  
Martin Meyer<sup>1,2,3</sup>, and Lutz Jäncke<sup>1,2,3,4</sup>

## Abstract

■ In this study, we used high-density EEG to evaluate whether speech and music expertise has an influence on the categorization of expertise-related and unrelated sounds. With this purpose in mind, we compared the categorization of speech, music, and neutral sounds between professional musicians, simultaneous interpreters (SIs), and controls in response to morphed speech–noise, music–noise, and speech–music continua. Our hypothesis was that music and language expertise will strengthen the memory representations of prototypical sounds, which act as a perceptual magnet for morphed variants. This means that the prototype would “attract” variants. This so-called magnet effect should be manifested by an increased assignment of morphed items to the trained category by a reduced maximal slope of the psychometric function as well as by differential event-related brain

responses reflecting memory comparison processes (i.e., N400 and P600 responses). As a main result, we provide first evidence for a domain-specific behavioral bias of musicians and SIs toward the trained categories, namely music and speech. In addition, SIs showed a bias toward musical items, indicating that interpreting training has a generic influence on the cognitive representation of spectrotemporal signals with similar acoustic properties to speech sounds. Notably, EEG measurements revealed clear distinct N400 and P600 responses to both prototypical and ambiguous items between the three groups at anterior, central, and posterior scalp sites. These differential N400 and P600 responses represent synchronous activity occurring across widely distributed brain networks and indicate a dynamical recruitment of memory processes that vary as a function of training and expertise. ■

## INTRODUCTION

In the last two decades, a vast number of studies have documented the profound influence of music (Elmer, Hänggi, Meyer, & Jäncke, 2013; Elmer, Meyer, & Jancke, 2012; Bermudez, Lerch, Evans, & Zatorre, 2009; Luders, Gaser, Jancke, & Schlaug, 2004; Keenan, Thangaraj, Halpern, & Schlaug, 2001; Schlaug, Jancke, Huang, & Steinmetz, 1995) and language (Ressel et al., 2012; Golestani, Price, & Scott, 2011) training, as revealed by professional musicians, early bilinguals, phoneticians, and simultaneous interpreters (SIs), on the functional and structural architecture of auditory-related brain regions. However, such plastic changes are often not restricted to single brain compartments, but rather affect a vast amount of cortical tissue (James et al., 2013; Zou et al., 2012; Bermudez et al., 2009; Abutalebi & Green, 2007; Gaser & Schlaug, 2003a, 2003b). This phenomenon probably accounts for the often observed cognitive advantages of music and lan-

guage experts in a variety of cognitive domains, including verbal learning (Bradley, King, & Hernandez, 2013; Kühnis, Elmer, Meyer, & Jancke, 2013), memory (Morales, Calvo, & Bialystok, 2013; Kraus, Strait, & Parbery-Clark, 2012; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011), attention (Strait, Kraus, Parbery-Clark, & Ashley, 2010; Costa, Hernandez, & Sebastian-Galles, 2008; Bialystok, Craik, Klein, & Viswanathan, 2004), and inhibition (Festman, Rodriguez-Fornells, & Munte, 2010; Bialystok et al., 2004).

Previous work on language development in childhood has proposed that language discrimination and expertise develop through mechanisms in which the prototype of a category acts as a perceptual magnet for all other category members (Kuhl, 2004). This becomes visible, for example, in that infants show a perceptual magnet effect for their native vowel category. In this context, American infants perceptually group the American vowel variants together but treat the Swedish vowels as less unified and more likely different (Kuhl, 2004). By contrast, Swedish infants show a reversed pattern in that they perceptually group the Swedish variants more than the American vowel stimuli (Kuhl, 2004). Such a perceptual magnet effect cannot only be observed in infants, but likewise in adults (Kuhl, 1991). Interestingly, some authors have argued that the magnet effect emerges from a simple similarity metric

<sup>1</sup>University of Zurich, <sup>2</sup>Center for Integrative Human Physiology, Zurich, Switzerland, <sup>3</sup>International Normal Aging and Plasticity Imaging Center, Zurich, Switzerland, <sup>4</sup>King Abdulaziz University, Jeddah, Saudi Arabia

\*These authors equally contributed to the study.

operating on collections of exemplars stored in memory, without the need to refer to special exemplars (Lacerda, 1995). On the basis of this reasoning, it is plausible to assume that music and language experts have much more multifaceted and detailed memory representations for single instrumental tones or phonemes compared with nonexperts. In fact, professional musicians are daily confronted with a wide heterogeneity of musical signals, especially when playing in orchestral ensembles. In a similar way, depending on the biological and biographical attributes of the speakers, professional SIs are trained daily to extract and recognize constant acoustic cues (i.e., phonemes) from very different speech signals. Exactly these multifaceted memory representations of experts are expected to strengthen the perceptual magnet effect toward the category of expertise, namely in the direction of speech or musical sounds.

Currently, only three neuroimaging studies (Specht, Osnes, & Hugdahl, 2009; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009; Husain et al., 2006) and one EEG study (Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2010) addressed the cognitive and neurophysiological correlates of categorization in laymen. In a first study, Specht and colleagues (2009) used a dynamic sound morphing paradigm in which white noise was gradually transformed into either speech or musical sounds and observed that the left STS responded almost solely to speech sounds (i.e., natural and morphed items), irrespective of the physical properties of the signals. By contrast, right-hemisphere homologues were equally responsive to the manipulation of both speech and musical items. In a second study, Staeren and colleagues (2009) presented to the participants a wide range of acoustic items, namely sounds of cats, female singers, acoustic guitars, and tones, and reported that distributed neuronal populations residing along the bilateral supratemporal plane were sensitive to categorical representations of sounds. In a third study, Husain and co-workers (2006) compared brain responses to a category discrimination task with an auditory discrimination task by presenting stimuli varying along a speech–nonspeech dimension as well as along a fast–slow temporal dynamics dimension. As a main result, the authors provided evidence for the contribution of frontal and parietal brain regions to auditory categorization. Finally, by presenting typical and ambiguous sounds from different categories (i.e., wood, metal, and glass), Aramaki et al. (2010) identified robust anterior and posterior N400 and P600 responses as objective markers of the cognitive mechanisms of auditory categorization.

Previous work has shown that N400 and P600 responses can reliably be elicited during sound categorization tasks, consisting in assigning prototypical and ambiguous items to the respective categories (Aramaki et al., 2010). The N400 can be described as a negative going brain response that develops at approximately 400 msec after stimulus onset at central-posterior (Kutas & Federmeier, 2011) or even anterior (Key, Dove, & Maguire, 2005) scalp sites.

So far, an impressive number of studies took advantage of the N400 component for investigating memory and categorization processes across different domains, for example, during language processing (Elmer, Meyer, & Jancke, 2010; Federmeier, McLennan, De Ochoa, & Kutas, 2002; Kutas & Hillyard, 1984), object, face, action, and gesture processing, as well as music processing (Elmer, Sollberger, Meyer, & Jäncke, 2013; Koelsch, 2011; Painter & Koelsch, 2011; Steinbeis & Koelsch, 2011). Meanwhile, it is assumed that the amplitude of the N400 response increases as a function of the memory requirements necessary for accomplishing a certain task (Federmeier et al., 2002).

A second brain response that has repeatedly been associated with cognitive abilities in general (Swaab, Brown, & Hagoort, 1998; Vanpetten, Kutas, Kluender, Mitchiner, & McIsaac, 1991) and with the allocation of memory resources in particular (Chung, Tong, & McBride-Chang, 2012; Ohara, Lenz, & Zhou, 2006) is called P600. The P600 is a positive going waveform peaking at around 500–600 msec poststimulus onset over central-posterior scalp sites (Friedman & Johnson, 2000). This specific ERP often co-occurs with the anterior N400 (Key et al., 2005) and is typically observed in the context of recognition tasks and memory recall paradigms (Kutas & Federmeier, 2011; Friedman & Johnson, 2000). Notably, the P600 has been shown to reflect explicit recognition memory (Friedman & Johnson, 2000), recollection (Olichney et al., 2000), successful retrieval, structural reanalysis, and repair functions across a wide range of stimulus material as well as semantic and episodic memory functions (Friedman & Johnson, 2000). Interestingly, P600 responses are also observed when participants encounter improbable or unexpected stimuli (Coulson, King, & Kutas, 1998).

Here, we used high-density EEG in association with a sound morphing paradigm (speech–noise [SN], music–noise [MN], and speech–music [SM] continua) to compare categorization processes between professional musicians, SIs, and controls. We predicted that music and language training will specifically influence the categorization of music and speech stimuli by showing a behavioral bias toward the category of expertise. This behavioral bias should therefore (1) result in a more frequent assignment of items from the middle part of the morphed continuum to the trained category, (2) result in a reduced maximal steepness of the psychometric function, and (3) be manifested in latency bands reflecting memory comparisons processes (i.e., N400 and P600 responses) in the time range from 300 to 1000 msec.

## METHODS

### Participants

Ten professional musicians (nine men; mean age = 35.1 years,  $SD = 7.9$  years; mean age of practice commencement = 7.4 years,  $SD = 1.7$  years; estimated total

number of practice hours since childhood = 25,079.6 hr,  $SD = 12,563.2$  hr), 10 professional SIs without formal musical education (nine women; mean age = 39.4 years,  $SD = 12.4$  years; mean years of practice = 17.4 years,  $SD = 12.4$  years), and 10 controls (6 women; mean age = 35.5 years,  $SD = 9.87$  years) participated in this study. All musicians (eight guitarists [four of them playing electric guitar], one string player, one bassist; primary musical instruments) commenced their musical training before the age of 11 years and were recruited from local music academies. All participants were consistent right-handers, as revealed by the Annett handedness inventory (Annett, 1970). Participants reported no current or past neurological, psychiatric, or neuropsychological diseases nor consumed illegal or legal medication. The local ethics committee approved the study, participants were paid for participation, and written informed consent was obtained from all participants.

### Cognitive Capability

To rule out differences in cognitive capability between the three groups, we adopted the KAI (Kurztest für allgemeine Basisgrößen der Informationsverarbeitung) and MWT (Mehrfachwahl-Wortschatz Intelligenz Test, MWT-B) tests. The KAI test permits to estimate the actual cognitive capability (fluid intelligence) and is based on working memory and speed of information processing. During this test, the participants had to read aloud meaningless sequences of 20 letters as quickly as possible as well as to repeat aloud auditory presented letters and digits increasing in length (up to nine items). The MWT test consists of 37 items, which are ordered as a function of difficulty level. For each item, the participants have to choose the unique word with a meaning out of five distractors (i.e., pseudowords). This procedure permits to estimate the crystalline intelligence of the participants in a short time and correlates fairly well with global IQ ( $r \sim .70$ ).

### Musical Aptitudes

To examine developmental and stabilized musical aptitudes as well as music achievement, all participants performed the AMMA (Advanced Measures of Music Audition) test (Gordon, 1989). This test consists of 30 successive trials in which participants compared pairs of piano melodies and then decided whether the melodies were equivalent (i.e., the exactly same acoustic pattern), rhythmically different, or tonally different. The test provides separate scores for rhythmic and tonal aptitude.

### Nonmanipulated Stimuli (Prototypical Items)

For this study, we selected three nonmanipulated stimuli that served as sources for parametric manipulations (morphing). These nonmanipulated stimuli were com-

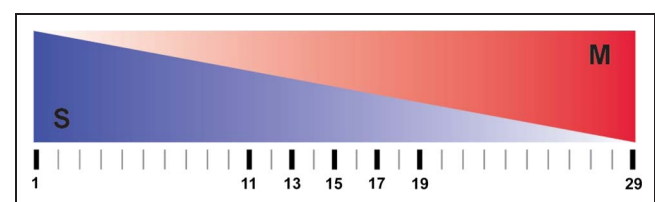
posed of the consonant–vowel syllable /ka/ (Elmer, Hänggi, et al., 2013; Elmer et al., 2012; Jancke, Wustenberg, Scheich, & Heinze, 2002), a c-major guitar chord (Fender Stratocaster, [www.samplitude.com/](http://www.samplitude.com/)), and pink noise (created with Adobe Audition, [www.adobe.com/](http://www.adobe.com/)).

### Parametrically Manipulated Stimuli (Morphing)

Starting from the three nonmanipulated stimuli, 29 linear transition steps between SN, MN, and SM were created by gradually morphing pitch, energy, spectrum, and rhythm parameters of the two respective signals (Figure 1). All parametric manipulations were performed by using the “Metamorpher” software ([www.peter-zorn.de/](http://www.peter-zorn.de/)). Furthermore, to smooth the transitions between the stimuli, the envelope of the consonant–vowel syllable was convolved to the guitar tone and to the noise signal by using Praat ([www.praat.org](http://www.praat.org)). In a similar way, the envelope of the guitar chord was convolved to the noise stimulus. Finally, five stimuli per category (i.e., SN, MN, and SM) were selected from the middle part of the morphed continuum (i.e., Steps 11, 13, 15, 17, 19; Figures 1 and 2; Specht et al., 2009) and presented to the participants together with the nonmorphed stimuli. All auditory items had a duration of 330 msec, were adjusted with a logarithmic fade-in/fade-out of 25 msec, recorded with a sampling rate of 44,100 Hz (16 Bit, mono files), and matched in mean intensity by using the Adobe Audition software ([www.adobe.com/](http://www.adobe.com/)). All auditory stimuli we used in the present work are available at the following link: [www.neuroscienze.ch/index.php?option=com\\_content&view=article&id=74&Itemid=79](http://www.neuroscienze.ch/index.php?option=com_content&view=article&id=74&Itemid=79).

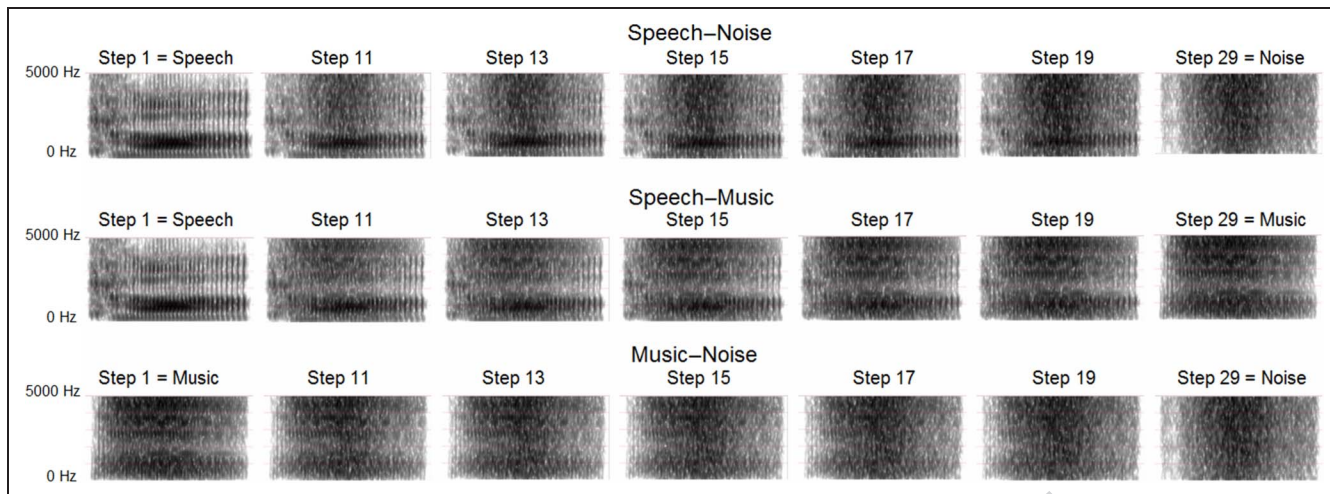
### Experimental Procedure

During EEG measurements, the participants were placed on a comfortable chair in a dimmed and acoustically shielded room, at a distance of about 110 cm from a monitor on which a fixation cross was presented. Before starting the main EEG session, the stimuli were shortly presented to the participants to gain familiarity with the acoustic material. During the main EEG session, participants were instructed to perform a categorization task consisting in assigning the presented auditory stimuli to the category



**Figure 1.** Morphed continuum between speech (S) and music (M). S1 and S29 depict prototypical items, whereas S11, S13, S15, S17, and S19 represent parametric morphing steps from the middle part of a SM continuum.





**Figure 2.** Spectrograms of the prototypical (Steps 1 and 29) and morphed (Steps 11–19) items for each condition (first row = SN, second row = SM, third row = MN) in the range from 0 to 5000 Hz ( $y$  axis).

of speech, music, or noise by pressing the respective response button with their right fore, middle, and ring finger. The auditory stimuli were presented via HIFI headphones (Sennheiser CX271, 70 dB sound pressure level) in a randomized order across four runs (14-min duration per run). Each of the 21 auditory stimuli (7 items  $\times$  3 conditions) was presented 20 times per run with a constant ISI of 1670 msec, resulting in a total number of 80 presentations per item. The presentation of the stimuli and recording of responses were controlled by the Presentation software (Neurobehavioral Systems, USA, [www.neurobs.com](http://www.neurobs.com)).

### EEG Recording and Analyses

A high-density EEG (128 channels) was recorded at a sampling rate of 250 Hz by using a band pass filter from 0.1 to 100 Hz (Electrical Geodesics, Eugene, OR). Electrode Cz was used as online reference, and impedances were kept below 30 k $\Omega$ . Before data preprocessing, the electrodes in the outermost circumference were removed, and noisy channels interpolated, resulting in a standard 109-channel electrode array. The data were low-pass filtered at 40 Hz, eye movement artifacts were corrected by using an independent component analysis (Jung et al., 2000), and remaining artifacts were removed manually by using the Brain Vision Analyzer software package (Brain Vision Analyzer 2.0; [www.brainproducts.com/downloads.php](http://www.brainproducts.com/downloads.php)). All electrodes were re-referenced to a virtual average reference, the data segmented into 2200 msec epochs, and a baseline correction was applied to the  $-200$  to  $0$  msec prestimulus time period. For each participant, item, and condition (i.e., SN, MN, and SM), epochs were averaged to calculate ERPs, and grand averages were computed for examining brain topologies. The averages for each stimulus, condition, and participant were exported and further analyzed in MATLAB ([www.mathworks.ch/products/matlab/](http://www.mathworks.ch/products/matlab/)) by using the Threshold-Free Cluster Enhancement

(TFCE) approach, which corrects for multiple comparisons ( $p < .01$ ). An important aspect of this approach is that it takes into account both data point's statistical intensity (here the amplitude of a single electrode) and the neighborhood (here the neighboring electrodes). Using this combined information reveals a more powerful comparison between groups and conditions especially for large data sets with large numbers of electrodes. A detailed description of this procedure can be found elsewhere (Mensen & Khatami, 2013).

### Behavioral and Biographical Data

To objectivize categorization of speech and music as a function of expertise and to avoid a double coding of responses (i.e., the total number of assignments to the two respective categories), we evaluated the relative number of speech (in the SN and SM conditions) and music (in the MN condition) responses in the range from 0 to 100%. In addition, between-group differences in categorization were evaluated for each participant, condition, and group by detecting the interval of maximal slope of the psychometric function across the morphed continuum. Here, it is important to mention that the maximal slope was only calculated between the single items situated in the middle part of the continuum (i.e., between S11, S13, S15, S17, and S19). The slopes between the prototypical items (i.e., S1 and S29) and the two morphed stimuli in the neighborhood (i.e., S11 and S19) were not evaluated, as these values are not informative at all for evaluating categorization differences. All slope values were quantified by calculating difference scores between the quotients (percentage values) of contiguous items. Statistical inference of biographical and behavioral data was performed in SPSS ([www-01.ibm.com/software/ch/de/analytics/spss/](http://www-01.ibm.com/software/ch/de/analytics/spss/)) by using ANOVAs (repeated measures), Mann-Whitney  $U$  tests, and  $t$  tests. Relationships between biographical

and behavioral data were estimated by using correlations (Pearson's  $r$ , one-tailed).

## EEG Data

ERPs were exported for each participant, stimulus, and condition, and statistical inference was assessed by mixed-model ANOVAs (corrected for multiple comparisons) in the context of the TFCE approach (Mensen & Khatami, 2013). The threshold for significance was set at  $p < .01$  (corrected for multiple comparisons). The ANOVAs were computed separately for the morphed ( $3 \times 5$  ANOVAs, 3 groups as between-subject factor and 5 morphed stimuli as repeated measurement factors) and nonmorphed stimuli ( $3 \times 3$  ANOVAs, 3 groups as between-subject factor and 3 nonmorphed stimuli as repeated measurement factors). Significant interaction effects were further evaluated by means of TFCE-based post hoc  $t$  tests.

## RESULTS

### Biographical and Behavioral Results

#### *Musical Aptitudes*

The three groups did not significantly differ in age,  $F(2, 9) = 0.673$ ,  $p = .433$ , nor in cognitive capability (MWT  $F(2, 9) = 0.735$ ,  $p = .414$ ; KAI  $F(2, 9) = 0.003$ ,  $p = .957$ ). The evaluation of the AMMA test ( $3 \times 2$  ANOVA, 3 groups as between-subject factor and 2 subtests as repeated measurement factors) yielded a main effect of group,  $F(2, 27) = 9.550$ ,  $p = .001$ , as well as a main effect of subtests,  $F(2, 27) = 14.555$ ,  $p = .001$ . All participants performed better on the rhythmical part of the AMMA test,  $t(29) = 3.861$ ,  $p = .001$ , one-tailed, the musicians performed better than SIs (tonal  $t(18) = 2.153$ ,  $p = .022$ ; rhythmical  $t(18) = 1.854$ ,  $p = .04$ ; one-tailed) and controls (tonal  $t(18) = 4.115$ ,  $p = .001$ ; rhythmical  $t(18) = 4.611$ ,  $p = .001$ ; one-tailed) on both subtests, and the SIs performed better than the controls in the rhythmical part,  $t(18) = 2.429$ ,  $p = .013$ , one-tailed. The latter result is possibly associated with the notion that professional language training may improve the analysis and memorization of rhythmical acoustic information (Christoffels, de Groot, & Kroll, 2006). We also evaluate relationships between the total number of estimated hours of musical training and behavioral performance in the AMMA test within the group of musicians (Pearson's correlations, one-tailed). This correlative analysis yielded a significant positive relationship between the total number of hours of musical training and behavioral performance in the AMMA test ( $r = .682$ ,  $p = .015$ ).

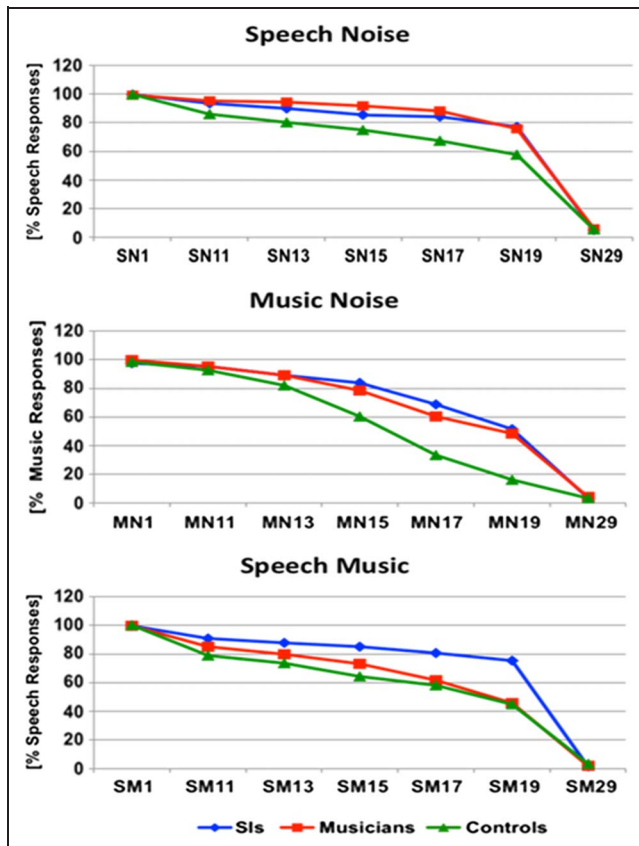
#### *Behavioral Responses*

Statistical comparison of the total number of correct responses to the nonmorphed speech, music, and noise

stimuli ( $3 \times 3$  ANOVA, 3 groups as between-subject factor and 3 items as repeated measurement factor) did not reveal a main effect of group,  $F(2, 27) = 2.303$ ,  $p = .141$ , nor a significant Group  $\times$  Item interaction effect,  $F(2, 27) = 1.318$ ,  $p = .284$ . Hence, these results attest a comparable assignment of the prototypical items to the respective category across the three groups.

The behavioral responses (percentage values) to the morphed transition steps were evaluated by performing separate  $3 \times 5$  ANOVAs (3 groups as between-subject factor and 5 morphed transition steps as repeated measurement factor) for each condition. This statistical procedure yielded significant Group  $\times$  Stimulus interaction effects in the MN,  $F(2, 27) = 3.735$ ,  $p = .037$ , and SM conditions,  $F(2, 27) = 3.859$ ,  $p = .034$ , but not in the SN condition,  $F(2, 27) = 1.043$ ,  $p = .366$ . Post hoc  $t$  test ( $t$  tests for independent samples, one-tailed) performed to disentangle these significant interaction effects revealed that the categorical assignment of MN stimuli significantly differed between musicians and controls (MN15  $t(18) = -1.762$ ,  $p = .047$ ; MN17  $t(18) = -1.781$ ,  $p = .046$ ; MN19  $t(18) = -2.175$ ,  $p = .021$ ) and SIs and controls (MN15  $t(18) = -2.472$ ,  $p = .012$ ; MN17  $t(18) = -2.667$ ,  $p = .008$ ; MN19  $t(18) = -2.849$ ,  $p = .005$ ). All between-group differences originated from a behavioral bias of the experts toward the category of music. In addition, during the SM condition SIs more often categorized the item SM19 as speech than controls,  $t(18) = -2.126$ ,  $p = .024$ , and musicians,  $t(18) = -2.119$ ,  $p = .024$ , did.

In an additional analysis, we subjected the maximal slope values of the psychometric functions (between two adjacent items) to one-way ANOVAs (3 groups as independent variable) separately for each condition. These analyses yielded a main effect of group for the MN,  $F(2, 27) = 5.149$ ,  $p = .049$ , and SM conditions,  $F(2, 27) = 12.627$ ,  $p = .006$ . The main effects of group originated from a less steep maximal slope in musicians,  $t(18) = -3.045$ ,  $p = .003$ , one-tailed, and SIs,  $t(18) = -2.621$ ,  $p = .008$ , one-tailed, during the MN condition compared with controls. Furthermore, during the SM condition, SIs showed a less steep maximal slope than musicians,  $t(18) = -2.797$ ,  $p = .006$ , one-tailed, and controls,  $t(18) = -3.594$ ,  $p = .001$ , one-tailed. An additional coding of the position of maximal steepness of the psychometric function within the morphed continuum revealed that SIs showed maximal steepness between the items MN15 and MN17, whereas in control participants the point of maximal steepness was situated earlier, between the items MN13 and MN15 (Mann-Whitney  $U = 21$ ,  $p = .019$ ). All other non-parametric comparisons did not reach significance. Taken together, these results provide clear evidence for different cognitive representations of speech and musical items as a function of expertise. Figure 3 shows typical psychometric functions for each group and condition and indicates that the attribution of the morphed transition steps to the respective categories increases as a function of the vicinity to the prototypical items.



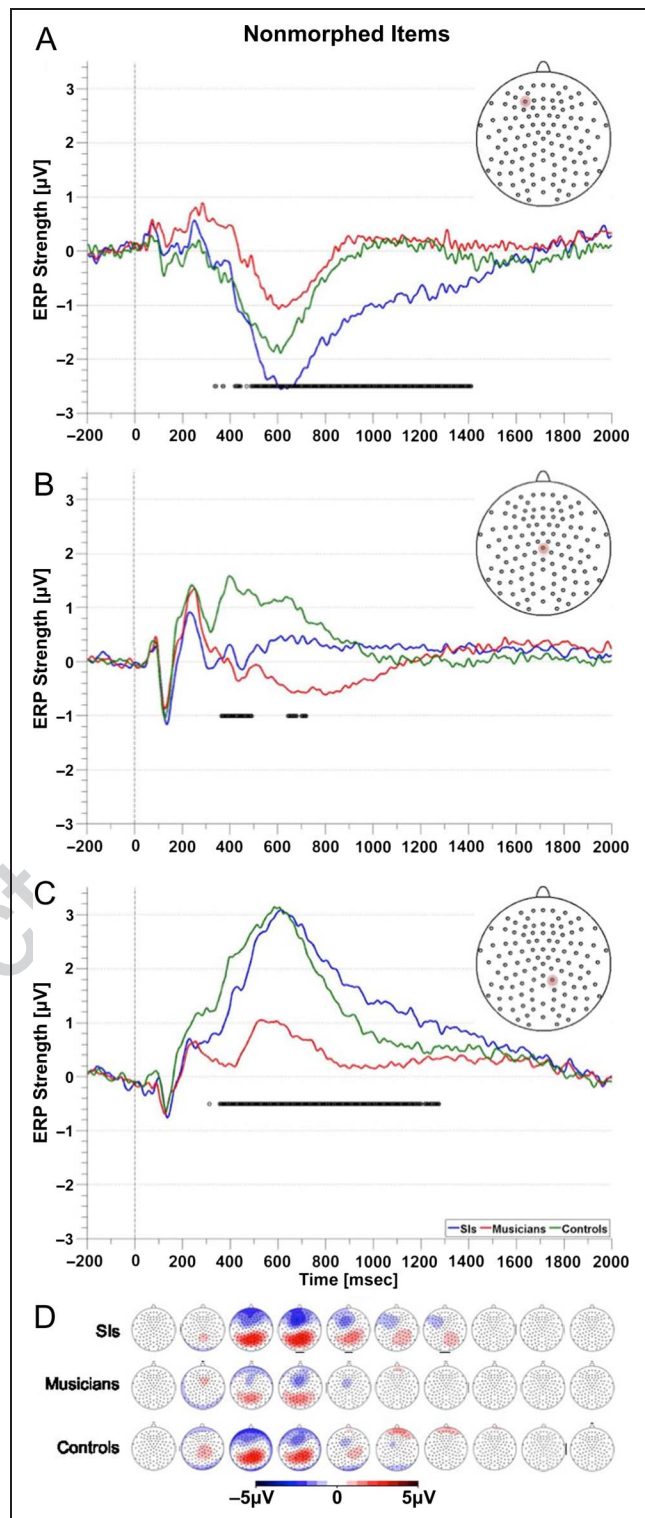
**Figure 3.** Psychometric functions for each condition and group. The y axis depicts the relative number (i.e., percentage values) of speech (in the SN and SM conditions) or music (in the MN condition) responses, the x axis the prototypical (S1 and S29) and morphed (S11–S19) stimuli.

## Electrophysiological Results

### Nonmorphed Stimuli

The whole-scalp ERP time courses evoked by the non-morphed stimuli were subjected to a  $3 \times 3$  ANOVA (3 groups  $\times$  3 stimuli) in the context of the TFCE approach. This analysis revealed a main effect of group,  $F(2, 27) = 27.002$ ,  $p = .0008$ , and stimulus,  $F(2, 27) = 124.699$ ,  $p = .0004$  (data not shown), as well as a significant Stimulus  $\times$  Group interaction,  $F(2, 27) = 10.737$ ,  $p = .007$ . These ERP effects were associated with three significant clusters of electrodes with maxima overall distributed along anterior (9 electrodes), central (Cz), and central-posterior (81 electrodes) scalp sites. The main effect of stimulus became particularly evident at central scalp sites in the time range from 100 to 900 msec poststimulus onset and was associated with different ERP morphologies as a function of the physical properties of the items. In the time range of the N1 component music and speech elicited an increased negativity compared with pink noise, whereas speech elicited increased P2 amplitudes compared with noise and music.

The main effect of group was most pronounced at anterior, central, and central-posterior scalp sites and mani-



**Figure 4.** Main effect of group during the nonmorphed conditions (speech, music, and noise) for the three single-subject populations, namely SIs (blue), musicians (red), and control group (green). (A) EEG data on the anterior peak electrode, (B) the Cz electrode, and (C) the central-posterior peak electrode. Electrode position is highlighted in red on the respective electrode map. On the y axis, ERP strength is shown in microvolts, the x axis depicts time in msec. (D) Topographic maps, averaged for 200 msec time windows, of the single groups, which are plotted according to the time axis. Significant group differences are highlighted by the black bars in the graph ( $p < .01$ ).



tested by clear distinct brain responses between the three groups in the time range between 350-1400 msec (Figure 4). At central-posterior electrodes, all groups elicited a positive going deflection resembling a P600 response between 300 and 1300 msec. Thereby, SIs and controls showed significantly increased amplitudes compared with musicians. Notably, at central scalp sites (i.e., at electrode Cz), all groups revealed an N400-like voltage devolution in the time range between 300 and 500. However, in the control group, this negativity was not as evident as in musicians and SIs. In addition, in musicians, the duration of the N400-like component was prolonged until 1000 msec poststimulus onset, whereas the controls and the SIs showed a completely reversed response pattern in the time range between 500 and 1000 msec, namely a positive deflection. This positivity was much more prominent in control participants compared with SIs. At anterior scalp sites, all three groups showed a huge negativity in the time range between 400 and 1400 msec. Thereby, the SIs were characterized by the most negative amplitudes, followed by the control participants. The musicians elicited the smallest amplitudes.

Finally, whole-head post hoc *t* tests for independent samples revealed that the Group  $\times$  Stimulus interaction effect originated from distinct waveforms between the three groups in response to speech, music, and noise

stimuli at anterior, central, and posterior scalp sites between 300 and 1300 msec. These post hoc analyses yielded highly significant group differences. However, the results were quite diffuse and heterogeneous. For the readers' convenience, all post hoc *t* tests are summarized in Table 1.

### Morphed Stimuli

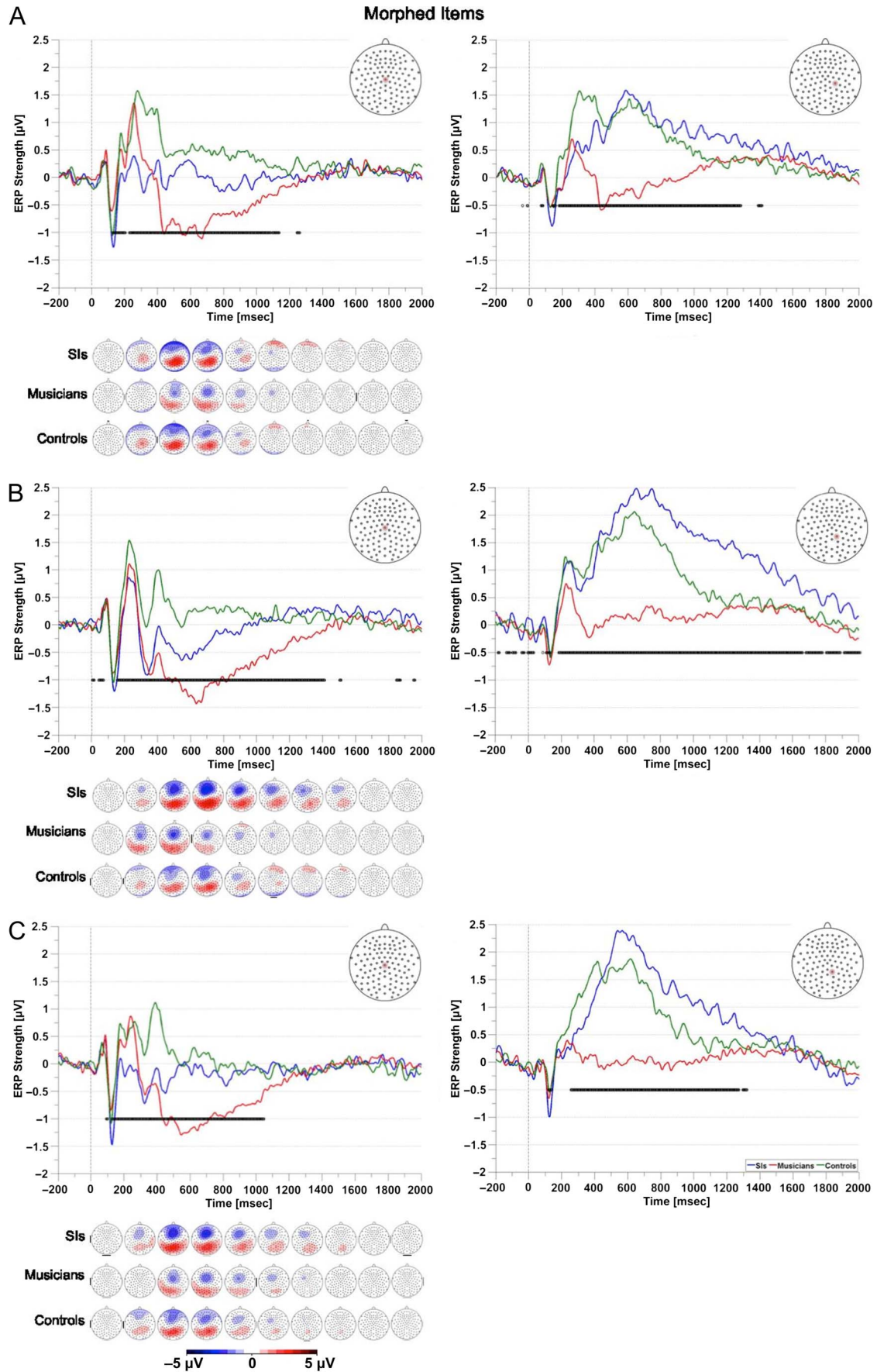
*SN condition.* The  $3 \times 5$  ANOVA yielded a main effects of group,  $F(2, 27) = 53.084, p = .00004$ , and stimulus,  $F(2, 27) = 9.068, p = .0008$ , with two significant clusters (maxima) situated at central (Cz) and right-central central (109 electrodes) sites. The main effect of group was manifested by clear distinct ERP morphologies between the three groups in latency bands ranging from 120 to 1300 msec. In particular, SIs showed an increased N1 response, whereas controls were characterized by increased P2 amplitudes. Furthermore, musicians showed an N400-like ERP at right-central scalp sites, whereas in the same time range controls and SIs more likely elicited a positive going P600 response (Figure 5). At electrode Cz, only musicians and controls showed an N400-like response pattern, whereas control participants were characterized by a P600 waveform. Interestingly, the N400 amplitude was much more increased in musicians

**Table 1.** Overview of the Post hoc Comparisons

Contrast	Condition	Scalp Location	Waveforms	Latency Bands (msec)	<i>p</i>
SI vs. M	Music	Central (Cz)	SI $\uparrow$ N400	300–400	.0027
		Central-anterior	SI $\uparrow$ P200	300–400	.0027
	Speech	Central-right	SI $\uparrow$ P600	500–800	.0027
	Noise	Anterior	SI $\uparrow$ N400	700–1000	.0008
		Posterior	SI $\uparrow$ P600	550–1300	.0008
SI vs. NM	Music	Central (Cz)	SI P600; NM N400	800–940	.0103
		Anterior	SI $\uparrow$ N400	650–1100	.0103
		Posterior	SI $\uparrow$ P600	500–1300	.0103
	Speech	–	–	–	<i>ns</i>
	Noise	Anterior	SI N400; NM P600	800–1300	.0103
		Posterior	SI $\uparrow$ N400	1000–1200	.0103
M vs. NM	Music	Anterior	M N400; NM P600	700–1300	.0075
		Posterior	NM $\uparrow$ N400	1000–1200	.0075
	Speech	Central (Cz)	M $\uparrow$ N400	400–850	.0004
		Anterior	NM $\uparrow$ N400	350–450	.0004
		Posterior	NM $\uparrow$ P600	300–850	.0004
	Noise	Central (Cz)	M $\uparrow$ N400	400–500	.0047
		Posterior	NM $\uparrow$ P600	500–900	.0047

$\uparrow$  = increased amplitudes.





compared with SIs. Finally, the main effect of stimulus developed in latency bands from 250 to 500 msec and reached its maximum at central scalp sites.

**MN condition.** The ANOVAs only revealed a main effect of group,  $F(2, 27) = 75.052, p = .0004$ , in the time range between 150 and 1800 msec, with two significant clusters (maxima) situated at central (Cz) and central-posterior (109 electrodes) scalp sites. In the time range of the N1 component (200–300 msec), the musicians elicited increased amplitudes than SIs and controls at central-posterior scalp sites. Furthermore, at the same electrode SIs and controls, but not musicians, elicited a positive going deflection resembling a P600 response between 300 and 1800 msec. The P600 was significantly increased in SIs compared with musicians. By contrast, at the same scalp location, the musicians more likely elicited an N400-like brain response. Finally, at central scalp sites especially musicians and SIs showed an N400-like deflection between 300 and 1000 msec. In the control participants, a similar negative deflection was not properly distinguishable, restricted to the time range 400–600 msec, and followed by a positive going ERP. The amplitude of the N400 deflection was increased in musicians compared with SIs, and the SIs significantly differed from the controls (Figure 5).

**SM condition.** ANOVAs revealed a main effect of group,  $F(2, 27) = 41.3, p = .0004$ , as well as a significant stimulus effect,  $F(2, 27) = 13.813, p = .0004$ , in the time range between 100 and 1000 msec, with maxima at central (Cz) and central-posterior (109 electrodes) scalp sites. At central electrodes, SIs showed an increased N1 and reduced P2 response compared with the two other groups. Furthermore, at central-posterior scalp sites, only SIs and controls showed a huge positivity resembling a P600 waveform in the time range between 150 and 1300 msec. Thereby, the amplitudes of this response were significantly increased in the SIs compared with the controls. A similar positivity was not distinguishable at all in the musician group. Interestingly, in the SM condition, all participants showed a negative going deflection in the time range between 400 and 1000 msec at central scalp sites (Figure 5). The main effect of group was associated with significant different N400-like amplitudes between the three groups, musicians showing most negative N400 values, and controls most positive ones. SIs were situated between musicians and controls. The main effect of stimulus was most prominent at central scalp sites and originated from differential physical attributes of the stimuli.

## DISCUSSION

### General Discussion

To the best of our knowledge, this is the first work that attempted to systematically investigate the influence of music and language expertise on the categorization of speech and musical items. Here, we provide first behavioral and electrophysiological differences between the three groups while categorizing speech and musical sounds as a function of expertise. In line with our hypotheses, the behavioral bias of the experts was manifested (1) by an increased assignment of items situated in the middle part of the morphed continuum to the trained category (i.e., speech or music) and (2) by a reduced maximal slope of the psychometric function within the morphed continuum of the expertise-related condition. In this context, musicians were characterized by a bias toward the category of music during the MN condition, whereas SIs revealed a behavioral bias toward the category of speech during the SM condition. Interestingly, during the MN condition, SIs were characterized by a similar devolution of the psychometric function as musicians (Figure 3, middle plot)—this result leading to suggest that language training has an influence on the cognitive representation of musical items as well. These behavioral results were accompanied by clear distinct N400 and P600 responses between the three groups over all experimental conditions at anterior, central, and central-posterior scalp sites. In this context, we provide strong evidence for different task-related processing modes of the experts' brain. In fact, the evaluation of the morphed transition steps yielded a consistent main effect of group. Taken together, our results indicate that speech and music training may influence the cognitive representations of speech and musical items. In turn, we will discuss the results in more details by integrating behavioral and electrophysiological data.

### Behavioral Results

In this study, we did not reveal behavioral differences between the three groups during the SN condition. This negative outcome is not really surprising, and different phenomena can be taken into account to explain why the psychometric function of SIs (i.e., the language experts) did not differ from the other two groups. First of all, signals with similar acoustic properties as pink noise occur widely in natural physical systems, and participants are generally experienced in extracting speech cues from noisy acoustic environments, as provided, for example, by the cocktail party phenomenon (Bronkhorst, 2000)

**Figure 5.** The main effect of group is depicted separately for each morphed condition. A = m speech noise, B = music noise, and C = speech music. SIs = blue, musicians = red, and controls = green. Electrode positions of the presented EEG data, Cz (left column) and peak electrodes (right column), are highlighted in red. Significant differences ( $p < .01$ ) are depicted as black bars in the graphs. Averaged topographic maps are averaged for 200 msec time frames and plotted under the respective time windows of the  $x$  axis. The  $y$  axis depicts ERP strength in microvolts.

or by hearing speech in noise (Parbery-Clark, Strait, Anderson, Hittner, & Kraus, 2011). This reasoning is consistent with the behavioral data pointing to a smoothed devolution of the psychometric function during the two speech conditions in comparison to the MN condition, at least at the end of the morphed continuum (Figure 3). Furthermore, the devolution of the psychometric function clearly reveals that within the SN continuum speech is generally perceived to be a particularly salient signal. Consequently, during this specific condition, speech signals are fairly robust to alterations of categorization as a function of training. An alternative explanation is that, in the context of temporally fluctuating noise, listeners generally experience a sort of “release from masking,” namely a better speech identification in fluctuating (i.e., pink) than in stationary (i.e., white) noise (Licklider & Miller, 1948). This specific acoustic phenomenon is called “listening in the valley” and can be explained by the notion that participants are able to extract speech information from the end of morphed signals, where the energy of the fluctuating pink noise reaches its energetic minimum (Ziegler, Pech-Georgel, George, Alario, & Lorenzi, 2005). In fact, by considering the spectrograms in Figure 2, it becomes apparent that power density falls at the end of the pink noise signal—this acoustic feature constituting a fundamental requirement for “listening in the valley.” Certainly, the same reasoning may even be true for musical items embedded in pink noise. However, music, unlike speech, has no any communicative relevance and is probably more likely perceived and categorized by focusing on aesthetic components, such as harmony, consonance, or dissonance (Besson & Schon, 2001). Hence, although music embedded in pink noise can saliently be detected at the end of the stimulus as well (i.e., like speech, “listening in the valley”), depending on the degree of noise in the signal, the cognitive and aesthetic representations of musical items may vary more strongly than those of speech as a function of repetitive exposure and training.

As a first main result, we provide behavioral evidence for an influence of expertise on the categorization of musical items during the MN condition. This expertise-dependent behavioral bias toward the category of music was manifested by a less steep maximal slope of the psychometric function in both musicians and SIs compared with controls. In addition, both expert groups more often assigned the items MN15, MN17, and MN19 to the domain of music than the control participants did. These results lead to suggest that musicians as well as SIs have much more robust memory representations for musical items, this cognitive function enabling them to recognize and categorize salient and prototypical aspects of musical cues even under aversive listening conditions. A self-evident explanation for this effect is that professional musicians are daily confronted with a wide heterogeneity of musical signals, especially when playing in orchestral ensembles. In a similar way, depending on the biological

and biographical attributes of the speakers, professional SIs are daily trained to extract and recognize constant acoustic cues (i.e., phonemes) from very different speech signals. Certainly, the same may be true for human beings in general. However, previous work has consistently shown that professional SIs more strongly engage cognitive resources like memory and attention functions (Elmer, 2013; Elmer, Hanggi, Meyer, & Jancke, 2011; Elmer, Meyer, Marrama, & Jancke, 2011; Cowan, 2010; Christoffels et al., 2006; Rinne et al., 2000) than other people do in a daily conversational context. Hence, we may assume that the behavioral bias of SIs toward the category of music has been driven, at least partially, by an increased engagement of memory resources (i.e., working memory and semantic memory; Rinne et al., 2000), by a more efficient allocation of attentive functions (Elmer, Meyer, et al., 2011), or even by an interaction between these two variables. Finally, an alternative explanation for the influence of interpreting training on the categorization of MN stimuli is the acoustic similarity between speech and musical signals. In fact, speech and music share many physical commonalities in that both signals convey acoustic information by means of timing, pitch, and timbre cues (Kraus & Chandrasekaran, 2010). Furthermore, perceptual and cognitive representations of speech and music overlap (at least partially) within the brain (Patel, 2011)—this spatial overlap facilitating transfer effects between the two domains (Elmer et al., 2012; Patel, 2011). Nevertheless, this argumentation can only be definitively proved or disproved by experimental paradigms specifically designed to investigate similarities and dissimilarities of speech and musical representations in the brain of musicians and language experts.

As a second main result, we revealed an influence of interpreting training on the categorization of SM stimuli. In fact, SIs more often categorized an item situated at the music end of the SM continuum as speech (SM19, mean = 73.31%) than musicians (SM19, mean = 45.78%) and control participants (SM19, mean = 44.8%) did. This behavioral bias was accompanied by a less steep maximal slope of the psychometric function in SIs compared with the other two groups. By considering Figure 3, it becomes visible that in both speech conditions (i.e., SN, upper plot; SM, lower plot) maximal steepness of the psychometric function arose between the stimuli S17 and S19, irrespective of group affiliation. By contrast, during the MN condition, maximal steepness developed earlier. This observation is important, in that it accentuates the affinity of human beings for speech signals. With this contextual framework in mind, our results are of noticeable relevance, in that they suggest that only intensive language training has the potential to further increase the already strong preference of human beings for speech sounds. In other words, professional language training enables to alter stimulus categorization in a domain in which human beings are generally experts.



## Electrophysiological Results

### *Nonmorphed Stimuli*

In line with the work of Aramaki and coworkers (2010), in this study, all three groups elicited robust anterior N400 as well as central-posterior P600 responses. Most notably, the magnitude of these two brain responses was clearly reduced in musicians, whereas SIs showed the most prominent negative deflection at anterior scalp locations. In addition, at central scalp sites (i.e., at electrode Cz), musicians and SIs elicited significantly increased N400 responses than control participants, in latency bands ranging from 300 to 500 msec. However, within the group of musicians, this negative deflection was prolonged until 1000 msec poststimulus onset. By contrast, in the time range from 500 to 1000 msec, SIs and controls showed a reversed response pattern, namely a positive deflection.

Our electrophysiological results clearly evidence that music and language expertise have a profound influence on the neuronal representation of prototypical items in widely distributed brain networks. Meanwhile, it is generally acknowledged that reduced ERP amplitudes reflect a minor recruitment (or less synchronicity) of neuronal ensembles. Therefore, the dampened N400 and P600 responses we revealed in musicians at anterior and posterior scalp sites are interpreted as representing a reduced cognitive load while categorizing prototypical items. Since during the task participants had to maintain the presented stimuli in STM and to compare them with the respective representations stored in long-term memory, our results are interpreted as indicating a less demanding or more efficient activation of prototypical items within distributed memory systems. This reasoning is further supported by previous studies showing that reduced N400 responses at anterior scalp sites are linked to a less demanding engagement of working memory resources (Vos, Gunter, Kolk, & Mulder, 2001; Mecklinger, Kramer, & Strayer, 1992) and semantic (Kutas & Federmeier, 2011) processes. In addition, posterior P600 magnitudes have repeatedly been shown to dramatically increase as a function of memory load, effort, and task demands (Friedman & Johnson, 2000).

A further interesting result of this study is that SIs showed substantially increased anterior N400 amplitudes in comparison with the other two groups. A similar result has previously been reported by Elmer et al. (2010), who provided first electrophysiological evidence for increased N400 responses in SIs compared with multilingual participants while performing a lexical decision task within and across languages. In addition, a very recent EEG study, which investigated conceptual memory associations (i.e., between tones, notes, and labels) in absolute pitch (AP) possessors and nonpossessors (NAP), clearly demonstrated that the better performance of AP musicians was achieved through an increased memory load, as reflected by enhanced N400 magnitudes (Elmer, Sollberger, et al., 2013). These two

previous EEG studies performed with music and language experts are important in that they lead to suggest that the increased frontal N400 responses of SIs were induced by an additional engagement of cognitive resources while performing the categorization task (Elmer, 2013). This is supported by a number of behavioral (Cowan, 2010; Christoffels et al., 2006) and neuroimaging studies (Elmer, Hanggi, et al., 2011; Elmer, Meyer, et al., 2011; Rinne et al., 2000), which have pointed out that SIs tend to maximize the engagement of cognitive control and memory mechanisms rather than to employ them in a parsimonious manner. It is likewise conceivable to assume that the differential electrophysiological response we revealed between musicians and SIs at anterior and posterior scalp sites may reflect two distinct training-dependent processing modes. In this context, two previous fMRI studies could show that in controls stimulus categorization is principally dependent on frontoparietal networks (Husain et al., 2006), whereas musicians more likely engage brain areas situated along the STS, at least during the categorization of spectrally complex sounds (Klein & Zatorre, 2011). In this context, we can only speculate on whether the prolonged maintenance of the N400 processing mode in musicians (at electrode Cz, in latency bands ranging from 500 to 1000 msec) may reflect a stronger recruitment of peri-sylvian brain regions. Certainly, although electrode Cz seems to most reliably represent brain activity originating from peri-sylvian areas (Baumann, Meyer, & Jancke, 2008), this speculative perspective needs to be further evaluated by studies combining EEG and fMRI.

### *Morphed Stimuli*

During the SN condition, SIs showed increased auditory-evoked N1 amplitudes, and control participants were characterized by enhanced P2 magnitudes. Because it is unreasonable to assume a perceptual encoding superiority of control participants, these early brain responses are interpreted as indicating an additional allocation of attentional resources to the morphed and therefore ambiguous SN items (Baumann et al., 2008; Hillyard, 1981; Picton & Hillyard, 1974). More interestingly, we revealed important morphological EEG differences between musicians and the other two groups at central scalp sites (i.e., central and right-central). Here, only the musicians showed a clear distinguishable N400 response, whereas SIs and controls more likely elicited a positive going deflection in the form of a P600 response. These distinct response patterns probably indicate different cognitive loads underlying the categorization of SN stimuli between musicians and controls. However, the specific cognitive processes reflected by this ERP are still a matter of debate (Kutas & Federmeier, 2011; Friedman & Johnson, 2000). In fact, it should be mentioned that, although the P600 waveform has repeatedly been associated with memory functions (Friedman & Johnson, 2000) and effortful aspects of cognitive processing (Elmer, Sollberger, et al., 2013; Friedman & Johnson,



2000; Swaab et al., 1998; Vanpetten et al., 1991), the same component can likewise be triggered when participants encounter improbable or unexpected items (Coulson et al., 1998). Because during the SN condition we did not reveal behavioral group differences, it is definitely conceivable that the increased P600 responses of SIs and control participants may not represent pure memory-related responses or differential cognitive loads, but possibly simply an effect of the participants' "surprise" upon encountering ambiguous speech items (Coulson et al., 1998). Otherwise, the missing P600 effect in musicians may be explained by the multifaceted acoustic variation they experience in everyday life.

During the MN condition, musicians elicited significantly increased N1 responses compared with the other two groups. Since a similar effect was not observed in response to the prototypical items, it is realistic to assume that this brain response reflects an additional allocation of attentional resources to the ambiguous musical items rather than an increased sensitivity of auditory-related brain regions. In addition, only SIs and controls elicited a P600 response at central-posterior scalp sites. By contrast, a similar waveform was not distinguishable at all in musicians. Most interestingly, only musicians and SIs elicited a clearly distinguishable N400 response at central scalp locations. Because musicians and SIs but not control participants showed a bias toward the category of music, it seems more plausible to assume that this behavioral bias was reflected by N400 rather than by P600 responses.

During the SM condition, SIs elicited significantly increased N1 and reduced P2 responses, indicating differential attentional requirements to the morphed items (Hillyard, 1981; Picton & Hillyard, 1974). Similar to the other two morphed conditions, the musicians elicited the strongest N400 responses, whereas SIs elicited the most prominent P600 waveforms. A previous EEG study (Aramaki et al., 2010) that investigated sound categorization clearly showed that the P600 response can be driven by conceptual associations between typical and ambiguous items. In addition, a very recent EEG study provided evidence for the fact that the unique faculty of AP possessors to associate tone with labels is best explained by their superiority in memory comparisons processes, as reflected by the P600 response (Elmer, Sollberger, et al., 2013). Most notably, this ERP was not present at all in NAP musicians. Finally, our results are also comparable with a previous work of Besson, Faita, and Requin (1994), showing that P600 responses were increased in experts (i.e., musicians) compared with nonexperts, but only when an explicit decision about the congruency of terminal notes was required from the participants. Taken together, our electrophysiological results provide evidence for a distinctive influence of speech and music expertise on task-related processing modes. These distinct processing modes were most probably driven by the engagement of different memory systems (i.e., working, episodic, and

semantic memory) and best reflected by N400 and P600 responses.

## Conclusions

In the present work, we combined behavioral and electrophysiological measurements for evaluating auditory categorization in professional musicians, SIs, and control participants. Our results are novel, in that we provide first evidence for a domain-specific behavioral bias as a function of training and expertise. Furthermore, we identified N400 and P600 responses as electrophysiological markers for group differences in categorization processes. However, our EEG data also accentuate that categorization is a dynamic cognitive process relying on widely distributed memory-related networks and cannot satisfactorily be described by focusing on single and partially overlapping ERP waveforms. Anyhow, our results contribute to a better comprehension of the still largely unexplored topic of categorization as a function of expertise.

## Limitations

A main limitation of the present work is that we can only speculate about the intrinsic meaning of the distinct P600 responses we revealed at central-posterior scalp sites as a function of expertise. In fact, the experimental paradigm we adopted in the present work does not permit to clearly associate this waveform with the specific cognitive mechanisms underlying categorization. Consequently, further studies applying oddball tasks in association with memory tasks may be helpful for better describing the specific cognitive processes reflected by the P600 component in experts and nonexperts. A further limitation of the present work is that gender was not counterbalanced across the three groups. Therefore, we cannot completely exclude that this variable may have influenced the data in some directions.

## Acknowledgments

This research was supported by the Swiss National Foundation (SNF Grants 320030-120661, 4-62341-05, and 320030B\_138668/1 to L. J.). We thank Nicolas Rüttiman for help in data acquisition. S. E. designed the experimental paradigm, formulated the hypotheses, evaluated the behavioral data, drafted the manuscript, and assisted C. K. with the evaluation of the EEG data. C. K. performed all EEG measurements and EEG data analyses and contributed to the manuscript. L. J. and M. M. contributed to the discussion and were involved in the preparation of the manuscript. J. K. was involved in the analysis of the behavioral data in association with psychometric functions. F. L. helped in preparing data for EEG analyses. In addition, C. K., J. K., and F. L. performed fruitful comments on a previous version of the manuscript. All authors read and approved the final version of this manuscript.

Reprint requests should be sent to Stefan Elmer, Division Neuropsychology, Institute of Psychology, University of Zurich, Binzmühlestrasse 14/25, CH-8050 Zurich, Switzerland, or via e-mail: s.elmer@psychologie.uzh.ch.

## REFERENCES

- Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, 20, 242–275.
- Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology*, 61, 303–321.
- Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., & Besson, M. (2010). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience*, 22, 2555–2569.
- Baumann, S., Meyer, M., & Jancke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience*, 20, 2238–2249.
- Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, 19, 1583–1596.
- Besson, M., & Schon, D. (2001). Comparison between language and music. *Biological Foundations of Music*, 930, 232–258.
- Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19, 290–303.
- Bradley, K. A. L., King, K. E., & Hernandez, A. E. (2013). Language experience differentiates prefrontal and subcortical activation of the cognitive control network in novel word learning. *Neuroimage*, 67, 101–110.
- Bronckhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica*, 86, 117–128.
- Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, 54, 324–345.
- Chung, K. K. H., Tong, X. H., & McBride-Chang, C. (2012). Evidence for a deficit in orthographic structure processing in Chinese developmental dyslexia: An event-related potential study. *Brain Research*, 1472, 20–31.
- Costa, A., Hernandez, M., & Sebastian-Galles, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, 106, 59–86.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21–58.
- Cowan, N. (2010). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting*, 5, 117–146.
- Elmer, S. (2013). The investigation of simultaneous interpreters as an alternative approach to address the signature of multilingual speech processing. *Zeitschrift für Neuropsychologie*, 23, 105–116.
- Elmer, S., Hanggi, J., Meyer, M., & Jancke, L. (2011). Differential language expertise related to white matter architecture in regions subserving sensory-motor coupling, articulation, and interhemispheric transfer. *Human Brain Mapping*, 32, 2064–2074.
- Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale facilitates the discrimination of temporal speech information in musicians. *Cortex*. doi.org/10.1016/j.cortex.2013.03.007.
- Elmer, S., Meyer, M., & Jancke, L. (2010). Simultaneous interpreters as a model for neuronal adaptation in the domain of language processing. *Brain Research*, 1317, 147–156.
- Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex*, 22, 650–658.
- Elmer, S., Meyer, M., Marrama, L., & Jancke, L. (2011). Intensive language training and attention modulate the involvement of frontoparietal regions during a non-verbal auditory discrimination task. *European Journal of Neuroscience*, 34, 165–175.
- Elmer, S., Sollberger, S., Meyer, M., & Jäncke, L. (2013). An empirical re-evaluation of absolute pitch: Behavioural and electrophysiological measurements. *Journal of Cognitive Neuroscience*, 25, 1736–1753.
- Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39, 133–146.
- Festman, J., Rodriguez-Fornells, A., & Munte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, 6.
- Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique*, 51, 6–28.
- Gaser, C., & Schlaug, G. (2003a). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, 23, 9240–9245.
- Gaser, C., & Schlaug, G. (2003b). Gray matter differences between musicians and nonmusicians. *Neurosciences and Music*, 999, 514–517.
- Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *Journal of Neuroscience*, 31, 4213–4220.
- Gordon, E. E. (1989). *Manual for the advanced measures of music education*. Chicago: G.I.A. Publications, Inc.
- Hillyard, S. A. (1981). Selective auditory attention and early event-related potentials—A rejoinder. *Canadian Journal of Psychology*, 35, 159–174.
- Husain, F. T., Fromm, S. J., Pursley, R. H., Hosey, L. A., Braun, A. R., & Horwitz, B. (2006). Neural bases of categorization of simple speech and nonspeech sounds. *Human Brain Mapping*, 27, 636–651.
- James, C. E., Oechslin, M. S., Van De Ville, D., Hauert, C. A., Descloux, C., & Lazeyras, F. (2013). Musical training intensity yields opposite effects on grey matter density in cognitive versus sensorimotor networks. *Brain Structure and Function*. doi:10.1007/s00429-013-0504-z.
- Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, 15, 733–746.
- Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37, 163–178.
- Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage*, 14, 1402–1408.
- Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: An ERP primer. *Developmental Neuropsychology*, 27, 183–215.
- Klein, M. E., & Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, 49, 878–887.
- Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, 8, 89–105.

- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11, 599–605.
- Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Neurosciences and Music IV: Learning and Memory*, 1252, 100–107.
- Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93–107.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843.
- Kuhnisch, J., Elmer, S., Meyer, M., & Jancke, L. (2013). Musicianship boosts perceptual learning of pseudoword-chimeras: An electrophysiological approach. *Brain Topography*, 26, 110–125.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Lacerda, F. (1995). *The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory*. Stockholm.
- Licklider, J. C. R., & Miller, G. A. (1948). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 20, 593.
- Luders, E., Gaser, C., Jancke, L., & Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. *Neuroimage*, 22, 656–664.
- Mecklinger, A., Kramer, A. F., & Strayer, D. L. (1992). Event related potentials and EEG components in a semantic memory-search task. *Psychophysiology*, 29, 104–119.
- Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage*, 67, 111–118.
- Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114, 187–202.
- Ohara, S., Lenz, F., & Zhou, Y. D. (2006). Sequential neural processes of tactile-visual crossmodal working memory. *Neuroscience*, 139, 299–309.
- Olichney, J. M., Van Petten, C., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia—Electrophysiological measures of impaired and spared memory. *Brain*, 123, 1948–1963.
- Painter, J. G., & Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music. *Psychophysiology*, 48, 645–655.
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *Plos One*, 6.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, doi:10.3389/fpsyg.2011.00142.
- Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials .2. Effects of attention. *Electroencephalography and Clinical Neurophysiology*, 36, 191–199.
- Ressel, V., Pallier, C., Ventura-Campos, N., Diaz, B., Roessler, A., Avila, C., et al. (2012). An effect of bilingualism on the auditory cortex. *Journal of Neuroscience*, 32, 16597–16601.
- Rinne, J. O., Tammola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., et al. (2000). The translating brain: Cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, 294, 85–88.
- Schlaug, G., Jancke, L., Huang, Y. X., & Steinmetz, H. (1995). In-vivo evidence of structural brain asymmetry in musicians. *Science*, 267, 699–701.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, 32, 771–783.
- Specht, K., Osnes, B., & Hugdahl, K. (2009). Detection of differential speech-specific processes in the temporal lobe using fMRI and a dynamic “sound morphing” technique. *Human Brain Mapping*, 30, 3436–3444.
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, 19, 498–502.
- Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience*, 23, 604–621.
- Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, 261, 22–29.
- Swaab, T. Y., Brown, C., & Hagoort, P. (1998). Understanding ambiguous words in sentence contexts: Electrophysiological evidence for delayed contextual selection in Broca’s aphasia. *Neuropsychologia*, 36, 737–761.
- Vanpetten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, 3, 131–150.
- Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, 38, 41–63.
- Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences, U.S.A.*, 102, 14110–14115.
- Zou, L. J., Abutalebi, J., Zinszer, B., Yan, X., Shu, H., Peng, D. L., et al. (2012). Second language experience modulates functional brain network for the native language production in bimodal bilinguals. *Neuroimage*, 62, 1367–1375.